DOCUMENT RESUME

ED 403 891 IR 056 264

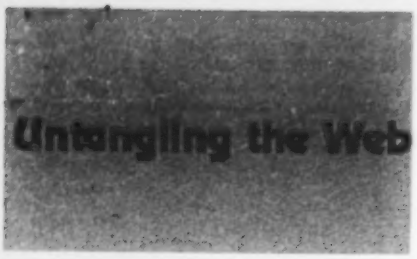
AUTHOR Tennant, Roy
TITLE The Art and Science of Web Server Management.
PUB DATE 96
NOTE 8p.; A downloaded version of one of 29 conference papers (for those in ERIC, see IR 056 258-274) placed on the Internet under the following overall title: "Untangling the Web. Proceedings of the Conference Sponsored by the Librarians Association of the University of California, Santa Barbara, and Friends of the UCSB Library" (Santa Barbara, CA, April 26, 1996).
AVAILABLE FROM Electronic version: http://www.library.ucsb.edu/untangle/
PUB TYPE Guides - Non-Classroom Use (055) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Networks; *Computer System Design; Computer Uses in Education; Error Correction; Higher Education; Hypermedia; *Information Management; Internet; *Online Systems; School Policy; Staff Development; Statistical Analysis; User Needs (Information); *World Wide Web
IDENTIFIERS Computer Users; *Links (Indexing); Technology Plans

ABSTRACT

Creating and maintaining a web server, particularly for a large organization, is both an art and a science. The science of link management, statistical analysis, and error redirection is only a part of what is required of web managers. Additional responsibilities include knowing the information needs of users; data owner support services; training others in web document design and markup; policy formation and enforcement; defining server standards; encouraging creativity; monitoring users; and checking links. This document describes the web manager's responsibilities, and provides instructions on how to customize default error messages and tips for web page referrals and redirection. No web manager is perfect, but those that approach perfection will bring a mix of skills and talents appropriate to the needs of the organization in building effective web systems. (Author/SWC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Roy Tennant
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

ED 403 891

The Art and Science of Web Server Management

Roy Tennant
Project Manager, Digital Library Research & Development
University of California, Berkeley

Copyright 1996, Roy Tennant. Used with permission.

Abstract

Creating and maintaining a web server, particularly for a large organization, is both an art and a science. The science of link management, statistical analysis, and error redirection is but a part of what is required of Web Managers. Additional responsibilities include defining server standards, monitoring users, training others in web document design and markup, policy formation and enforcement, and data owner support services. No Web Manager is perfect, but those that approach perfection will bring a mix of skills and talents appropriate to the needs of the organization in building effective web systems.

Introduction

The following pages describe the best advice I have for managers of world wide web servers, based upon my experience with two large web servers at an academic research library. These servers encompass thousands of files and hundreds of individual "data owners" -- those with the information being provided.

Implied by my advice is a vision of the perfect Web Manager. Since this is the case I feel I must be clear that this person does not exist. I certainly do not do everything that I advocate -- rather I strive every day to attain it. I must also admit that there is as much blind luck and witchcraft involved as there is "art and science". Most web managers learn on the job and make it up as they go along.

Much of what follows is general in nature; that which is not refers to NCSA httpd running under Unix, which is probably the most widely implemented web server at colleges and universities in the United States.

Know Your Clientele

The first precept of building any information system is to know your clientele and their information needs. This is complicated by the fact that you often have at least two clienteles -- your local users (which have your primary allegiance) and the entire Internet (which comes along "for free"). So how do you serve both?

IR056264

BEST COPY AVAILABLE

Often you can do both quite well at the same time. After all, your "local" users are in many cases "remote", in that they may never come into the library and yet have need of your services. Sometimes it is as easy as realizing you should include your area code with any phone numbers you provide. Other times it may be complicated by requirements to limit access to local users for certains kinds of information.

Knowing what information your clientele needs is but one part of the equation -- you also need to know how they will wish to use it. Structuring information for good usability on-screen is very different than doing so for printing out on paper. Occasionally you may even find it necessary to provide a second version of your information optimized for printing. Providing versions of your documents in "rich" formats such as Adobe Acrobat may also be required.

Your users will be accessing your information from a variety of hardware and software platforms. Luckily most web servers keep track of exactly what client software is being used to access your server. NCSA httpd keeps track of this information by inserting one line in a log file called "agent_log" for each client access of your server. For example:

```
Mozilla/2.0 (Win16; I)
```

This line tells you that someone accessed your server using the MS Windows 16-bit version of Netscape Navigator 2.0. The location of this file is often `/var/log/httpd/agent_log`.

Foster Your Data Owners

Those who have responsibility for the content being served on your web server are the "owners" of the information and therefore should be involved in how it is presented. This involvement could be as trivial as advising the person who marks it up in HTML or as substantive as doing it themselves. Those falling into the latter category should be supported in several ways:

Training
No one is born knowing HTML, so markup training is certainly a beginning requirement. But beyond that your data owners may require training in FTP, format conversion (e.g., Adobe Acrobat files, word processing to HTML conversion utilities), image acquisition hardware (e.g., scanners, digital cameras) and software (e.g., Adobe Photoshop), Internet searching techniques, and effective browser use.

Templates
Providing basic HTML templates for different types of common file structures is a very helpful way to save time creating new files. The simplest way to do this is to write an HTML file "shell" that data owners can then fill in with their own information (e.g., http://www.lib.berkeley.edu/Web/Templates/). For some types of files that do not require sophisticated markup you may want to program a script that will create a web document with the appropriate basic markup for your server simply by filling in a few blanks on a form (e.g., http://www.lib.berkeley.edu/Web/Forms/minimal.html). This technique can be particularly important to allow those who do not have the time or inclination to learn HTML to contribute information to your web server.

Advice
Chances are you do more HTML as a web manager than anyone else in your organization. Therefore you have likely learned a few things along the way. Share those tips and tricks with your data owners. Create a mailing list of your data owners so that sharing information with them as a group is easy.

Assistance
Although you may not have the time to hold everyone's hand, sometimes there is simply no substitute for sitting down with someone and showing them how to do something or even doing it for them while they are there to tell you how they want it.

Current Awareness

You are busy, but so are your data owners. If you see a new web resource that you think would be of interest to someone, forward the information on it to them. The majority of your data owners are unlikely to be monitoring any of the general-purpose current awareness resources such as NET-HAPPENINGS (http://www.mid.net:80/NET/).

Formulate Policies

Sooner or later you will need to write policies that govern appropriate uses of your web server. With a written policy that has been reviewed by appropriate groups within your organization you will have the organizational authority and support to refuse requests that fall outside of the policy. It can also be cited when necessary to stop inappropriate uses that you may discover. For examples of web policies, see Susan Brown's collection at http://www.cc.colorado.edu/Library/InfoSource/Current/wwwpol.html and Stacey Kimmel's at http://www.lib.muohio.edu/~skimmel/wwwpols.html .

Enforce Your Standards

A server without standards is likely to be a disaster. If you have not set up standards regarding "look and feel", for example, how will your users know when they are at your site and when they have left? If you allow your data owners to do anything they like they often will. Determine the essential limits on creativity to protect usability and user friendliness and stick to them. But also be careful not to set limits when none need exist. Reassess your standards periodically. If they are too strict, loosen up. If you are encountering problems where no standards exist or where they are too lax, then tighten your grip.

Encourage Creativity

The flip-side to dictating a common look and feel is encouraging creativity. The trick is to strike a reasonable balance. One way in which this can be done is to specify the presence of some simple graphic element or text that will place the page in context (such as a button bar that identifies the institution) and leave the rest to the data owner.

The Web Development Team at the UC Berkeley Library had originally tried to specify a particular structure for each page that described a Library branch. There are good reasons for doing that, but as the Web Manager I soon retreated from such a repressive position to encourage creativity. I was soon glad that I did, as data owners become much more engaged in the process of creating their own distinctive presence within the framework that we had provided. The end result has been an overall enrichment of the server, albeit at the cost of having a different structure for most branch pages.

Another result of encouraging creativity is that those who are the most creative set examples that inspire and challenge others. This kind of contagion can be a much more effective motivator than dictating participation.

Monitor Your Users

Most web servers keep exhaustive statistics on client requests. Logs include one that simply specifies the type of client software used (agent_log in NCSA httpd, see above), the kinds of errors they are experiencing (error_log), and what they are looking at (access_log). All of these logs are typically available in your `/var/log/httpd` directory.

A number of programs are available to report your access statistics in various ways. One of the best overviews of your options is "WWW Usage Statistics" at http://www.uiowa.edu/~libsci/studentalumni/asis/d-lib.htm .

Study Your Competition

A Web Manager should be on the web as much as possible every day. See what others are doing and how they are doing it. Copy their markup. Get graphic design ideas. Decide what you don't like by seeing the mistakes of others.

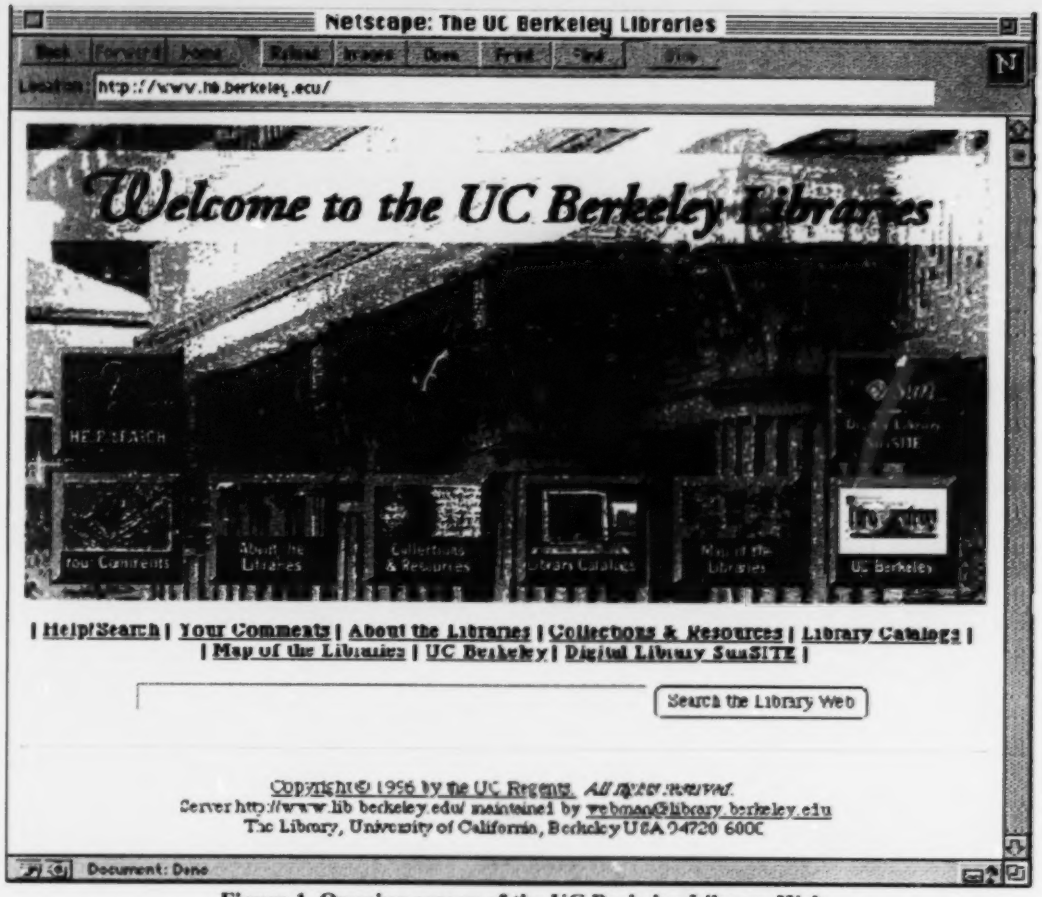

Netscape: The UC Berkeley Libraries
Location: http://www.lib.berkeley.edu/
Welcome to the UC Berkeley Libraries
| Help/Search | Your Comments | About the Libraries | Collections & Resources | Library Catalogs |
| Map of the Libraries | UC Berkeley | Digital Library SunSITE |
Search the Library Web
Copyright© 1996 by the UC Regents. All rights reserved.
Server http://www.lib.berkeley.edu/ maintained by webman@library.berkeley.edu
The Library, University of California, Berkeley USA 94720 6000
Document: Done

Figure 1. Opening screen of the UC Berkeley Library Web

An excellent resource for seeing what other libraries are doing on the web is Libweb, a directory of library-based web servers available at http://sunsite.berkeley.edu/Libweb/ .

Check Your Links

As all web managers know, a constant problem is keeping the links on your server up-to-date. Documents are moved every day, thereby making links go dead, and it is your responsibility as a Web Manager to identify help your data owners identify these dead links so they can be fixed. Luckily there is software available that automates this task, such as MOMspider (http://www.ics.uci.edu/WebSoft/MOMspider/).

You also need to make sure that people are successfully linking to information on your server, and this can be monitored by periodically reviewing the error log. The error log will tell you when someone tried to access something unsuccessfully. You can often spot problems this way that can be easily corrected once you become aware of them.

Strive to Do Better

No matter how good you are doing as a web manager, there is probably always room for improvement. Here are some ways that you can make sure you are doing as well as you could be, since they tend to be things that are overlooked by many web managers.

Error Message Replacement

BEST COPY AVAILABLE

Most web server error messages are cryptic at best and user hostile at worst. The default "file not found" error

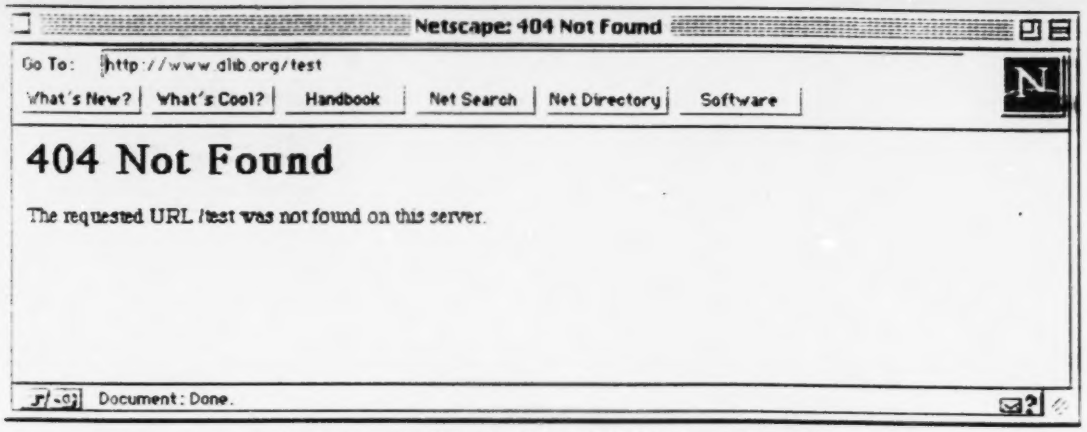

Netscape: 404 Not Found
Go To: http://www.dlib.org/test
What's New? What's Cool? Handbook Net Search Net Directory Software
404 Not Found
The requested URL /test was not found on this server.
Document: Done.

Figure 2. Default screen for a "404 Not Found" error.

This error message tells the user nothing about what they might try to do to find what they need. Use the built-in ability of most web servers to replace the standard error messages with something more useful:

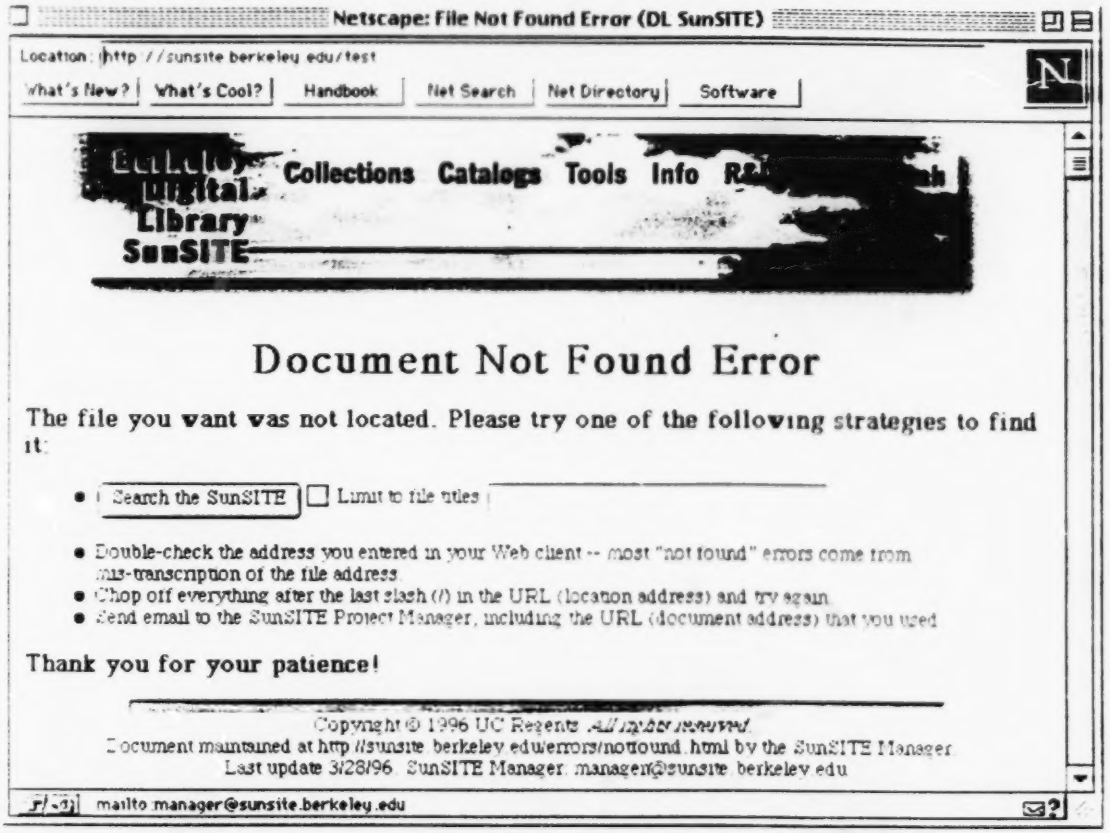

Netscape: File Not Found Error (DL SunSITE)
Location: http://sunsite.berkeley.edu/test
What's New? What's Cool? Handbook Net Search Net Directory Software
Collections Catalogs Tools Info
Digital Library SunSITE
Document Not Found Error
The file you want was not located. Please try one of the following strategies to find it:
Search the SunSITE
Limit to file titles
Double-check the address you entered in your Web client -- most "not found" errors come from mis-transcription of the file address.
Chop off everything after the last slash (/) in the URL (location address) and try again.
Send email to the SunSITE Project Manager, including the URL (document address) that you used.
Thank you for your patience!
Copyright © 1996 UC Regents. All rights reserved.
Document maintained at http://sunsite.berkeley.edu/errors/notfound.html by the SunSITE Manager.
Last update 3/28/96. SunSITE Manager: manager@sunsite.berkeley.edu
mailto:manager@sunsite.berkeley.edu

Figure 3. Screen of a replacement error message for a "404 Not Found" error.

It is relatively simple to implement this. For NCSA httpd servers there is a file called "srm.conf", in the "conf" subdirectory wherever your httpd directory is located on your server. In this configuration file (plain text), you will find:

```
# If you want to have files/scripts sent instead of the built-in version
# in case of errors, uncomment the following lines and set them as you
# will. Note: scripts must be able to be run as if the were called
# directly (in ScriptAlias directory, for instance)
```


BEST COPY AVAILABLE

```
# 302 - REDIRECT
# 400 - BAD_REQUEST
# 401 - AUTH_REQUIRED
# 403 - FORBIDDEN
# 404 - NOT_FOUND
# 500 - SERVER_ERROR
# 501 - NOT_IMPLEMENTED

ErrorDocument 404 /errors/notfound.html

#ErrorDocument 302 /cgi-bin/redirect.cgi
#ErrorDocument 500 /errors/server.html
#ErrorDocument 403 /errors/forbidden.html
```

Every line with a pound symbol (#) in front of it is a comment and is thus ignored by httpd. Therefore, to set up an HTML file to be used instead of the default error message all that must be done is to remove the pound symbol in front of the line referring to the error type and specify the address (URL) of the file to be substituted (e.g., the line above for ErrorDocument 404).

Once you have created your replacement error message file (probably placed in an "errors" subdirectory of your web server root) you will need to have your system administrator edit the "srm.conf" file for you and restart the server. After that your users are likely to be much happier.

Referrals

One of the most direct ways to prevent users from losing your information when you move it is to create a page that points to the new location. On such a page you will typically request that the user update any bookmarks or links that they have to the old location. For example:

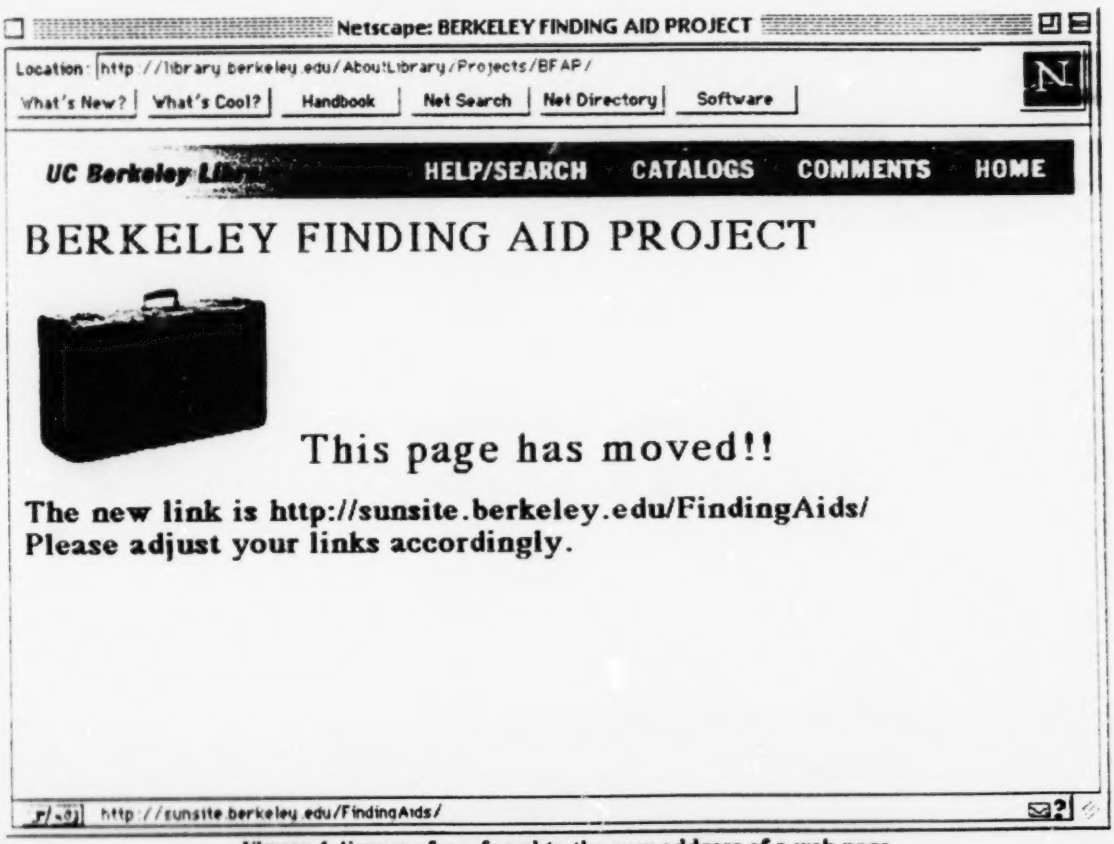


Figure 4. Screen of a referral to the new address of a web page.

BEST COPY AVAILABLE

Redirects

- Redirection is the procedure whereby you can specify the new location of a document behind the scenes. Then, when a user requests the document at the old location they are automatically redirected to the new location. This prevents the "404 File Not Found" error message as well as the need to put up an explicit referral as discussed above. Also, you can combine redirection with error message replacement so when a file is not found it is first checked against a list of moved files. If a new location is not found for the requested file then the user is presented with a helpful error message file. An example of a script that can handle this operation is "RedMan: WebPage Redirection Manager" available at http://sw.cse.bris.ac.uk/WebTools/redman.html .

Question Everything

It is all too easy to assume that your first decisions on the design of your server are the best. Struggle against the law of inertia (objects at rest tend to stay at rest) and periodically question all of your decisions. Ask yourself if there is a better way to organize your information or design a particular page. Ask for critical evaluations from your users and other staff in your organization.

Look critically at each web page for accuracy, usability, "scannability" (the ability to scan page and quickly discover your options), good visual design (is it engaging but not cluttered or distracting?), and good markup (does it display well in different browsers?)

The Perfect Web Manager

The perfect web manager will be part publicist, part politician, and part programmer. She will learn constantly, read voraciously, and advocate tirelessly. He will thrive in uncertainty and relish change. She will be a scientist, an artist, and a witch with a good luck streak.

Questions or comments on this piece may be sent to the author at rtennant@library.berkeley.edu . Prepared for the *Untangling the Web* Conference at the University of California, Santa Barbara, April 26, 1996.

Direct questions or comments about this page to webmaster@library.ucsb.edu .

Search "Untangling" Papers | "Untangling the Web" Home | InfoSurf Home

This is an official University of California, Santa Barbara Library web page.